**Further analysis**

All analysis will be implemented in R 3.3.2.

**Confounding effect of perceived closure**

Though it is commonly used, a concern with the paradigm employed is the confounding of predictability ratings with the perception of closure, where pitches that are perceived as poor endings to a phrase are also considered more surprising. In order to measure the potential effect of this confound, closure ratings were collected alongside ratings of expectedness.

First, the degree of correlation between expectancy and closure ratings will be calculated, collapsed across Time.

*cor.test(expectancy.ratings, closure.ratings)*

Second, linear mixed effects models with and without closure ratings as a covariate will be compared. Variance explained by closure ratings will give us an idea of how much the percept of closure influences the rating of surprise/expectedness.

*prediction <- lmer(prediction ~ Time + (1+melody.id|Participant), data = tidydata)*

*prediction.closure <- lmer(prediction ~ Time + Closure.ratings + (1+melody.id|Participant), data = tidydata)*

*anova(prediction, prediction.closure)*

*r2.difference <- r2.prediction – r2.prediction.closure*

**Compartmentalization of style**

The IDyOM configuration used in this study employs an LTM trained on either only tonal music (“Before”) or a combination of tonal and atonal music (“After”) in order to approximate the listeners’ musical exposure. However, given evidence of the compartmentalization of musical style, or the ability to generate different predictions based on the style listened to (e.g. Hansen, Vuust, & Pearce, 2016), we also would like to compare the information content and entropy produced by an LTM trained only on atonal music to an LTM trained on both tonal and atonal music. This applies to the “After” portion of the study alone.

The primary data pattern to provide evidence for compartmentalization would be if the information content and entropy of the atonal-only trained LTM correlated with human ratings better than the tonal+atonal-trained LTM. Overlap in the 95% confidence intervals of the two correlations would indicate there is no significant difference between the two types of LTM.

*Combined.LTM.cor.prediction <- cor.test(expectancy.ratings, combined.information.content)*

*Atonal.LTM.cor.prediction <- cor.test(expectancy.ratings, atonal.information.content)*

*Combined.LTM.cor.precision <- cor.test(certainty.ratings, combined.entropy)*

*Atonal.LTM.cor.precision <- cor.test(certainty.ratings, atonal.entropy)*